Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

A99.9
F764Un
Cop. 2

United States
Department of
Agriculture

Forest Service

**Intermountain
Forest and Range
Experiment Station**

Research Note
INT-317

August 1981

Formatting and Documenting Multidisciplinary Data

Joyce A. Schlieter[1]

ABSTRACT

The management of data accumulated during the course of a 5-year multidisciplinary research program is discussed. A system developed for organizing collected data is described and a reference sheet, designed to identify and facilitate data access, is shown.

KEYWORDS: data management, multidisciplinary data, data reporting form

The Intermountain Forest and Range Experiment Station Forest Residues Utilization Research and Development (R&D) Program was chartered for 5 years (1974-1979) to investigate: (1) opportunities for reducing the wood volume left as residue following harvest; (2) the biological and environmental consequences of alternative harvesting and intensive utilization practices; and (3) new or expanded markets for currently unused timber resources. Vast amounts of data were accumulated by Program scientists and numerous cooperating researchers as various parts of the above problems were studied.

Most data was generated by program research concerned with environmental and management consequences of harvesting. Some disciplines involved in this portion of the Program were meteorology, hydrology, nutrient cycling, microbiology, entomology, and silviculture. Analyses were carried out in each discipline to obtain first-order responses to harvesting. One of the goals of the R&D Program was to link all or parts of the data from the various disciplines in a fashion that would be directly useful to forest managers. To facilitate analysis and especially to aid in accessing and combining data from different disciplines, an efficient data management system was needed.

Efficient data handling, storage, and analysis required the use of a computer. The available computer system used punched cards as the input medium. An advantage of punched cards was that a permanent backup of the data base was established. The data cards were read into the computer and then were transferred to magnetic tape for storage.

DATA FORMAT

Early in the Program, the decision was made to use a common formatting system for individual disciplines. Each participating researcher was assigned a set of numbers and from these a unique three-digit format code was given to each type of data. For example, in meteorology some of the codes were:

050 --- air temperature
051 --- soil temperature
052 --- precipitation
053 --- solar radiation
054 --- net radiation

Much of the research effort for the Program was concentrated on a single experimental area. A grid of permanent plot points was set up for the area and all observations collected were tied to these permanent plots. A standard card format was then defined (fig. 1) with the first 14 columns set aside for identification. This common identification block makes it easy to cross-reference data from the different disciplines by date and site. A similar system was used for sampling at other experimental areas.

DATA DOCUMENTATION

The next step in the management of the data was to develop a system for documentation. This was important for several reasons. At the termination of the Program, some of the scientists involved either were moved to other locations or were assigned to other research projects. They were no longer immediately available for advice and consultation, even though continuing analysis and modeling work will be underway for some time. In addition, the data base accumulated during the Program's life contains a vast amount of information that might be of potential value to other researchers.

---

[1] The author is mathematical statistician stationed at the Intermountain Station's Forestry Sciences Laboratory, Missoula, Mont.

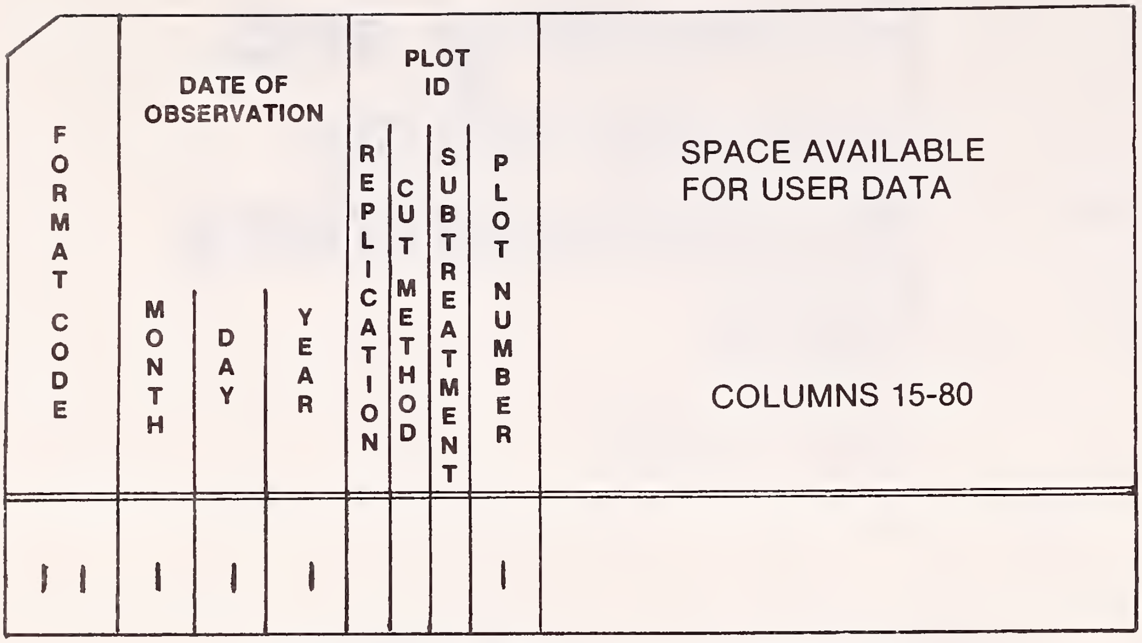

Figure 1.--Data card design for the Forest Residues R&D Program.

A data reporting form was designed to meet these documentation needs. Each researcher provided information on the sampling procedure used, the dates of sampling and the card layout (an example is given in fig. 2). The sheets were compiled in looseleaf notebooks under six categories: fauna, residues, soils, vegetation, meteorology, and hydrology. This reference system will aid any future user of the R&D data base.

LONG-TERM DATA STORAGE

Since the end of the R&D Program, an effort has been made to identify the data likely to be in demand for further analysis. These data are currently being stored on a nine-track, high density tape (6250 bits/inch) at the LBL Computer Center in Berkeley, Calif. This storage medium was chosen because of its reliability and accessibility. Data files not available on the tape are available on cards.

Each file on the master tape is identified by a pathname. These pathnames will be listed on the data reporting form. For example, the pathname for the data described in figure 2 would be:

JOYCE/DATA/METEOR/HUNGER/PRECIP/
CORAM/F 052

This pathname, the card layout information from the data reporting form, and the tape number are all one would need to access and use the data.

DISCUSSION AND RECOMMENDATIONS

This method for formatting and documenting multidisciplinary data was the approach taken by **one** research program. Data management and analysis requirements will rarely be the same for two research endeavors. In 1974 when the Program began, time, as well as personnel and available computer facilities, limited the options. If the Program were starting now, different methods might be employed.

The heavy dependence on cards for data entry proved to be costly and time consuming. In 1977, data acquisition equipment was installed on two experimental areas for recording meteorological measurements on cassette tape. The information on these cassette tapes was then transferred directly to magnetic tape and a computer routine was used to translate the data into the original format (fig. 1).

The use of an available data base management system is another option that would warrant examination. Most such systems are designed for business and accounting applications, but there have been attempts to handle research data (Anderson and Cohen 1976; Anderson 1977). An advantage to such a system would be that information could be retrieved without requiring the services of a computer programmer. Often there are lengthy delays in obtaining desired analyses due to a lack of programing assistance. A disadvantage of using an existing data management system would be the costly computer resources and personnel required for implementation. If the computer center being accessed had a system available, however, there would be no additional computer expense. Personnel specifically trained in data management should be available because experienced statisticians are not necessarily "experts" in the field of computerized data management (Helms 1978).

Documentation is the foundation of a sound data management system. Through the data reporting form an accurate and complete record of all the information collected can be maintained. The long-term data storage methods allow easy retrieval of the data files. These procedures designed and used by the R&D Program could be adapted to other similar research investigations.

FOREST RESIDUES R-D PROGRAM DATA REPORTING FORM

TYPE OF DATA: Meteorology (Precipitation) **FORMAT CODE:** 052

PRINCIPAL INVESTIGATOR: Roger Hungerford

LOCATION: Coram

GENERAL INFORMATION:

Tipping bucket rain gages that recorded events every 0.01 inch of moisture were used. The orifice is 8 inches in diameter. A 7-day drum chart recorder was used to record events. Rain gages were set on the ground surface or on a low stump within 25 feet of the weather shelter. Rain gages were set out in May each year and removed for the winter in November.

DATES OF SAMPLING:

Station:	12	14	21	23
Date Started:	6/13/75	9/20/73	8/18/74	6/1/74
Date Off:	Continuing	9/30/77	9/30/77	9/30/77

DATA FORMAT:

Columns	*Items*	*Columns*	*Items*	*Columns*	*Items*
1-3	Format code 052	39-42	Precip 1400 X.XX		
4-5	Month	43-46	Precip 1600 X.XX		
6-7	Day	47-50	Precip 1800 X.XX		
8-9	Year	51-54	Precip 2000 X.XX		
10	Replication	55-58	Precip 2200 X.XX		
11	Cut Method	59-62	Precip 2400 X.XX		
12	Subtreatment	64	Code*		
13-14	Plot Number	66-71	Last day missing data		
15-18	Precip 0200 X.XX				
19-22	Precip 0400 X.XX				
19-22	Precip 0400 X.XX				
23-26	Precip 0600 X.XX				
27-30	Precip 0800 X.XX				
31-34	Precip 1000 X.XX				
35-38	Precip 1200 X.XX				

*See code sheet at beginning of Meteorological data

Figure 2.--Sample data reporting form for the Forest Residues R&D Program.

PUBLICATIONS CITED

Anderson, Gary D.
1977. Software for management of research data. *In* Proc. Stat. Comput. Sec., Am. Stat. Assoc., Washington, D.C. p. 56-60.

Anderson, Gary D., and Eli Cohen.
1976. A language for management of non-rectangular case oriented data files for statistical analysis. *In* Proc. Stat. Comput. Sec., Am. Stat. Assoc., Washington, D.C. p. 58-67.

Helms, Ronald W.
1978. An overview of research data management. *In* Proc. Stat. Comp. Sec., Am. Stat. Assoc., Washington, D.C. p. 10-19.

The Intermountain Station, headquartered in Ogden, Utah, is one of eight regional experiment stations charged with providing scientific knowledge to help resource managers meet human needs and protect forest and range ecosystems.

The Intermountain Station includes the States of Montana, Idaho, Utah, Nevada, and western Wyoming. About 273 million acres, or 85 percent, of the land area in the Station territory are classified as forest and rangeland. These lands include grasslands, deserts, shrublands, alpine areas, and well-stocked forests. They supply fiber for forest industries; minerals for energy and industrial development; and water for domestic and industrial consumption. They also provide recreation opportunities for millions of visitors each year.

Field programs and research work units of the Station are maintained in:

- Boise, Idaho
- Bozeman, Montana (in cooperation with Montana State University)
- Logan, Utah (in cooperation with Utah State University)
- Missoula, Montana (in cooperation with the University of Montana)
- Moscow, Idaho (in cooperation with the University of Idaho)
- Provo, Utah (in cooperation with Brigham Young University)
- Reno, Nevada (in cooperation with the University of Nevada)

U.S.D.A.
NAT'L AGRIC LIBRARY
NOV 25 '91
CURRENT SERIAL RECORDS